

# Gestión y Aplicación de Metadatos Asociados al Tráfico Multimedia en Videoconferencia 3D

José M. Cubero Muñoz<sup>1</sup>, Antonia M. Sanz Fernández<sup>1</sup>, Enrique Estalayo Gutiérrez<sup>1</sup>, Pablo Pérez García<sup>2</sup>,  
Fernando Jaureguizar Núñez<sup>1</sup>, Julián Cabrera Quesada<sup>1</sup>, Jaime J. Ruiz Alonso<sup>2</sup>

{jcm, asf, eeg, fjn, jcq}@gti.ssr.upm.es

<sup>1</sup>Grupo de Tratamiento de Imágenes  
Universidad Politécnica de Madrid  
Ciudad Universitaria s/n  
28040 Madrid

{pablo.perez, jaime.ruiz}@alcatel-lucent.com

<sup>2</sup>Alcatel-Lucent  
c/María Tubau, 9  
28050 Madrid

**Resumen** — En los terminales multimedia de nueva generación se están incorporando cada vez más funcionalidades capaces de generar información adicional que enriquece la comunicación audiovisual en sí misma. Esta nueva información adicional, metainformación o metadatos, deberán poder ser gestionados y transmitidos de forma independiente pero sincronizados con el propio flujo multimedia. En este trabajo se presenta una arquitectura que permite resolver el problema identificado y que ha sido implementada con éxito en el proyecto CENIT VISION, dando soporte a diversos servicios de valor añadido asociados a una Videoconferencia 3D inmersiva: interacción con el usuario, grabación de sesiones, gestión de calidad, etc. La misma solución también tiene aplicación en otros entornos del sector multimedia como publicidad personalizada para IPTV, y de otros, como videovigilancia.

## I. INTRODUCCIÓN

En los entornos actuales de comunicación multimedia, tanto en comunicaciones personales como en difusión de contenidos, está tomando cada vez más relevancia la capacidad de enriquecer la pura transmisión audiovisual con información adicional que permita dotar al servicio de comunicaciones de mayor valor. Así, por ejemplo, las plataformas de televisión digital son capaces de proporcionar aplicaciones interactivas asociadas al contenido que se está mostrando en un momento determinado, gracias a la adopción de arquitecturas interactivas como MHP (*Multimedia Home Platform*). También las redes IMS (*IP Multimedia Subsystem*) y TISPAN (*Telecommunications and Internet converged Services and Protocols for Advanced Networking*), evolución del concepto de red inteligente, facilitan la provisión de servicios de comunicación avanzados.

A esta información relativa al contenido multimedia la denominaremos genéricamente metadatos (lit. “datos sobre los datos”). Éstos pueden referirse al propio contenido multimedia (metadatos descriptivos, p. ej., determinación del número de personas en la imagen), a su estructura y relación interna (metadatos estructurales, p. ej., clasificación en escenas o capítulos) o a información sobre su gestión (metadatos administrativos, p. ej., permisos de acceso al contenido) [1].

En un caso típico de contenido audiovisual, los metadatos se generan en la fase de producción. Así, en la grabación y montaje de un programa de televisión se emplean metadatos que ayudan a etiquetar las diferentes partes del programa y, por tanto, permiten su realización. Sin embargo, esta información rara vez está disponible una vez que el contenido ya ha sido producido (se rompe la cadena de metadatos). En otros servicios de comunicaciones, tipo videoconferencia, donde no existe ese proceso de producción, se carece de metadatos intrínsecamente asociados al proceso de captura. En la mayor parte de los escenarios reales se parte de un contenido multimedia “cerrado” (ya codificado, sincronizado y multiplexado) del que se obtiene metainformación desde una fuente paralela. Esta información después se describirá, transmitirá en un flujo independiente al contenido multimedia original y/o almacenará haciendo uso de formatos de codificación y gestión de metadatos como MPEG-7 [2] [3].

El presente trabajo describe la solución adoptada para la inserción, codificación, sincronización y transmisión de metadatos en un entorno de Videoconferencia 3D de última generación que ha sido desarrollado como parte del proyecto CENIT VISION [4].

En un contexto de este tipo, y el de VISION no es una excepción, la cadena de generación, codificación y sincronización del flujo multimedia es bastante compleja: obligar a que cualquier fuente de metadatos se introduzca de forma sincronizada como un elemento de esta cadena supone un coste de integración muy elevado. Por tanto, es preciso que el flujo de metadatos se genere de forma externa. Sin embargo, también es necesario que se transmita de forma sincronizada con el contenido multimedia, de modo que cualquier elemento de la red disponga de la metainformación necesaria para poder proporcionar servicios como los mencionados anteriormente. Si el flujo de metadatos viaja referenciado temporalmente con el flujo multimedia, el receptor será capaz de capturar ambos flujos y de utilizar los datos adicionales a posteriori (p.ej., al visualizar un contenido grabado). En la solución propuesta se ha obtenido una sincronización fina entre los metadatos y el contenido hasta el nivel de los cuadros de vídeo.

En la sección II se describe el sistema de gestión de metadatos desarrollado, que cumple con las características mencionadas en el párrafo anterior. En la sección III se discuten las aplicaciones de este sistema en diversos escenarios. Y, finalmente, la sección IV describe las principales conclusiones del trabajo.

## II. ARQUITECTURA DEL SISTEMA

En esta sección se describe la arquitectura de referencia que da soporte a la generación y sincronización de metadatos que permiten ofrecer diferentes servicios de valor añadido asociados a un determinado contenido audiovisual: una videoconferencia, publicidad, una película, etc. Esta arquitectura se representa en la Figura 1, mostrando los elementos que participan en una comunicación unidireccional, en este caso, orientada a una sesión de Videoconferencia 3D e implementada para el proyecto VISION.

Los requisitos que esta arquitectura satisface, sin un coste elevado asociado, son: la generación de metadatos, tanto genéricos como particulares para cada aplicación, permitiendo la adición de nuevos elementos generadores de metadatos e incluso, de aplicaciones que hagan uso de los mismos; y la sincronización de los metadatos con el contenido audiovisual.

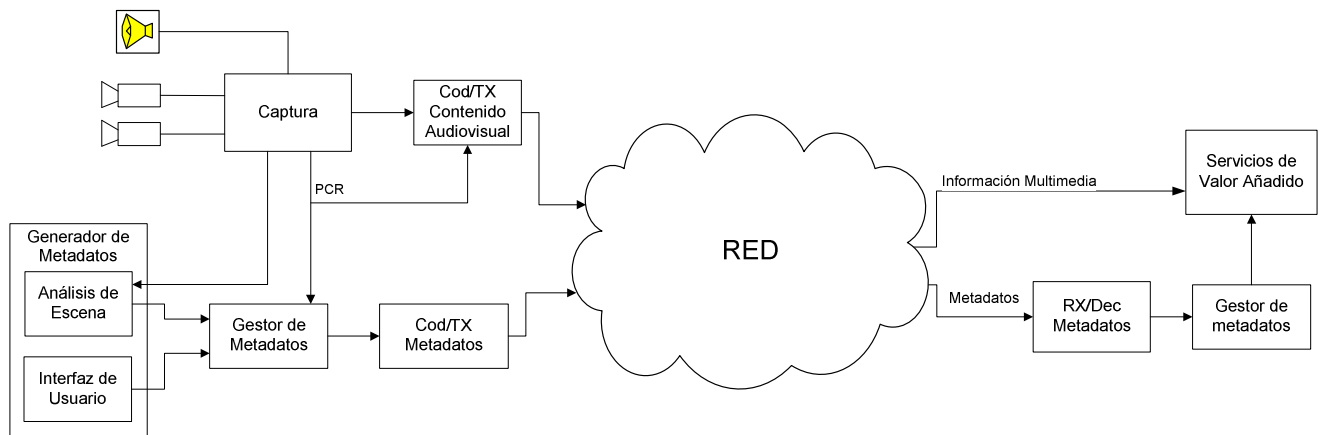


Figura 1. Arquitectura implicada en la gestión y aplicación de metadatos (unidireccional) desarrollada para el proyecto VISION

A continuación se describen los componentes que intervienen en la generación, gestión y aplicación de los metadatos utilizados para ofrecer servicios de valor añadido asociados a una Videoconferencia 3D:

- **Generador de Metadatos.** Este componente engloba a todos los elementos del sistema que detectan eventos y generan metadatos asociados a los mismos. En la arquitectura utilizada en VISION, los metadatos pueden producirse tanto por el interfaz de usuario como por el analizador de escenas.
  - **Módulo de Análisis de Escena.** Este elemento se encarga de analizar la escena y generar un metadato en función del evento detectado. Algunos ejemplos de eventos a detectar son: si el usuario está o no presente en la sala en la que se esté llevando a cabo la videoconferencia (información de presencia), si está manipulando algún tipo de objeto, si está caminando, etc.
  - **Interfaz de Usuario.** Este componente permite al usuario de la videoconferencia realizar acciones que se traducen en la generación de nuevos metadatos. Un ejemplo de estos metadatos son las etiquetas textuales que el usuario puede asociar a un instante o elemento de la escena.
- **Gestor de Metadatos.** En transmisión, este módulo recibe los metadatos procedentes de los diferentes generadores de metadatos existentes y les añade la marca de tiempo PCR (*Program Clock Reference*), procedente del Módulo de Captura, para que los metadatos estén sincronizados con el contenido audiovisual correspondiente a la videoconferencia. En recepción, el gestor de metadatos es el encargado de seleccionar y facilitar los metadatos que requiere el servicio de valor añadido correspondiente.
- **Módulo Codificador/Transmisor de Contenido Audiovisual.** Representa las funcionalidades de codificación y transmisión del contenido de audio y vídeo propio de la Videoconferencia 3D.
- **Módulo Codificador/Transmisor de Metadatos.** Este elemento se encarga de encapsular en un flujo TS (*Transport Stream*) los metadatos generados y transmitirlos hacia la red para que puedan ser utilizados por la aplicación final, en este caso la Videoconferencia 3D.
- **Receptor/Decodificador de Metadatos.** Este módulo recibe el flujo de metadatos procedente de otros usuarios, lo decodifica y se lo proporciona al Gestor de Metadatos para que éste, a su vez, lo transfiera a la aplicación final que hará uso de los mismos.
- **Servicios de Valor Añadido.** Este bloque representa a la aplicación final que hará uso de los metadatos que están sincronizados con el flujo de audio y vídeo. En la solución aplicada en VISION, la aplicación que hace uso de estos metadatos presta un servicio de Videoconferencia 3D, de modo que los metadatos se utilizan en servicios ofertados

al usuario final y se relacionan con la videoconferencia, como es el caso de la grabación de sesiones, o bien, se emplean para monitorizar el servicio mediante técnicas de QoS (*Quality of Service*).

El intercambio de información entre los diferentes componentes que conforman la arquitectura de VISION, y que aquí se ha descrito, se lleva a cabo mediante comunicaciones UDP, codificándose los diferentes metadatos en forma de mensajes de texto.

#### A. Generación de metadatos

En el proyecto VISION, se considera metadato a la información asociada a eventos que se producen durante el transcurso de una Videoconferencia 3D. Estos metadatos pueden ser generados por el propio usuario o por el módulo encargado de analizar la actividad que se produzca en la sala en la que se esté llevando a cabo dicha videoconferencia.

Los metadatos generados por el usuario, a través de su interfaz, consisten en este caso en información textual con la que se selecciona y clasifica una determinada escena. Por otra parte, el módulo de Análisis de Escenas genera y comunica los metadatos de eventos relacionados con la actividad del usuario, como por ejemplo: información de presencia, usuario paseando por la sala, número de personas en la sala, manipulación de objetos, etc.

En la arquitectura aplicada a la plataforma VISION (ver Figura 1), estos son los dos elementos encargados de generar metadatos. Sin embargo, el diseño del sistema permite añadir nuevos elementos capaces de producir metadatos. Basta con que estos elementos establezcan una conexión con el gestor de metadatos a través de una conexión UDP.

Un metadato utilizado en esta plataforma de videoconferencia 3D contiene la siguiente información básica en función del suceso al que esté asociado:

- Un identificador del evento que caracteriza de manera unívoca al evento con el que se asocia el metadato generado.
- Un identificador de usuario que relaciona al usuario con el evento de metadato. Este identificador será, en el caso de VISION, una dirección SIP.
- La marca de tiempo, en unidades de PCR, en la que se ha producido el evento para permitir la sincronización del metadato con el flujo de vídeo.
- Una prioridad para la manipulación del contenido audiovisual en función de los metadatos, como es la segmentación de secuencias, la recuperación de escenas, etc.
- Información complementaria al evento, texto introducido por el usuario al seleccionar una escena. Existen metadatos que no exigen este dato, como es el caso de los que contienen información de presencia.

#### B. Captación de metadatos

Tal y como se ha indicado, los metadatos son generados y transmitidos hacia la red para que diferentes aplicaciones puedan hacer uso de ellos. Es el Gestor de Metadatos el que decide qué metadatos deben transmitirse y hacia dónde.

A la hora de transmitir, el Gestor de Metadatos mantiene una conexión UDP con el Módulo de Análisis de Escena y otra similar con la Interfaz de Usuario (ver Figura 2). A través de estas conexiones el Gestor recibe los metadatos caracterizando los eventos detectados en la videoconferencia, que contienen la información descrita en el apartado anterior correspondiente al identificador del evento, información complementaria sobre el evento en caso de que sea necesaria e información sobre el usuario generador del evento. En función de esta información, el Gestor de Metadatos decide la prioridad que le corresponde a dicho metadato. Además, les añade una marca de tiempo en función del PCR recibido y le comunica toda esta información al Codificador y Transmisor de Metadatos, que es el encargado de transmitirlo hacia el receptor como un flujo de información independiente del contenido audiovisual.

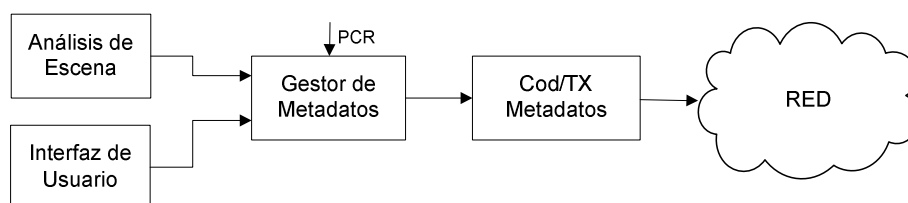


Figura 2. Arquitectura para transmisión de metadatos en la plataforma VISION

En recepción (Figura 3), de manera similar a transmisión, el Gestor de Metadatos establece una conexión UDP con el Receptor/Decodificador de Metadatos y con los Servicios de Valor Añadido que usarán dichos metadatos. En este caso, el Gestor de Metadatos recibe el flujo de metadatos sincronizado con el contenido audiovisual y facilita la información del metadato requerida en cada instante al Servicio de Valor Añadido de Videoconferencia 3D.

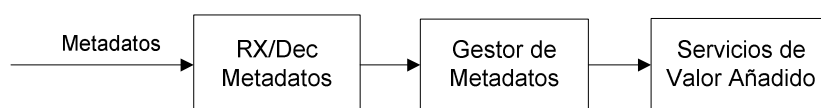


Figura 3. Arquitectura para recepción de metadatos en la plataforma VISION

### C. Inserción y sincronización de metadatos en el flujo multimedia

El módulo de Transmisión de Metadatos realiza la función de inserción y sincronización de metadatos. La necesidad de sincronización del flujo de metadatos con el flujo de información multimedia viene dada por la propia naturaleza de las transmisiones IP, que permiten encaminar independientemente diferentes flujos de información. De este modo, tanto los metadatos como la información multimedia pueden llevar rutas independientes, perdiéndose el sincronismo entre ambos flujos. Sin embargo, esta misma característica de las redes IP permite mantener intacta la cadena de producción y transmisión de contenidos.

El módulo presentado en esta sección permite referenciar el flujo de metadatos a la información multimedia (audio y vídeo), sincronizando ambos flujos de información. Así, se obtiene una transmisión sobre IP de un flujo de metadatos que se mantiene independiente de la preexistente cadena de producción y transmisión de contenidos, sin perder por ello la característica de sincronismo que se obtendría directamente si los metadatos se insertaran multiplexados con la información multimedia, modificando la cadena de producción.

La información de sincronización temporal puede ser procesada por etapas posteriores en la red o por aplicaciones finales de usuario. En el marco del trabajo presentado, el flujo de información multimedia está encapsulado en paquetes de transporte de MPEG-2 (MPEG-2 TS [5]) que se transmite opcionalmente dentro de una sesión RTP, este flujo de información junto con el flujo de metadatos insertado y sincronizado es utilizado por aplicaciones finales de usuario, como muestra en la subsección III.A.

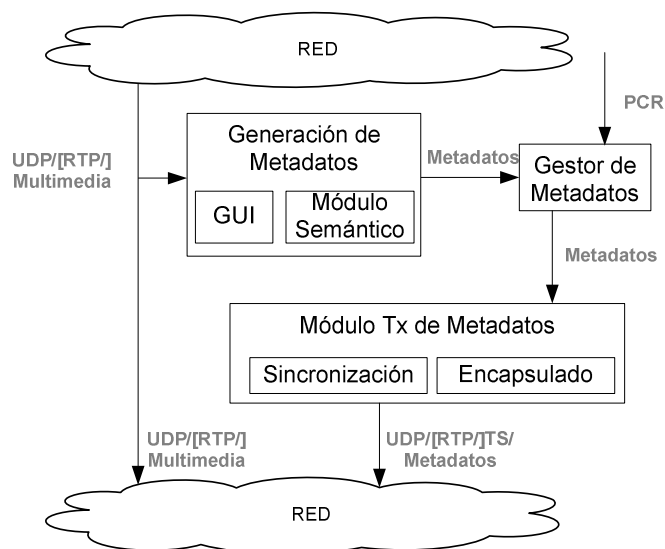


Figura 4. Módulo de Inserción y Sincronización de Metadatos.

La sincronización de los metadatos se hace efectiva añadiendo información temporal en el flujo de metadatos. Esa información temporal se obtiene en función del reloj del flujo multimedia en el instante de generación del metadato por el interfaz gráfico del usuario (GUI) o por el reconocedor semántico (Módulo Semántico en la Figura 4) del módulo de Análisis de Escena (ver subsección II.A.).

Estos módulos son los encargados de enviar la información extraída del reloj del flujo multimedia, en particular, el PCR (ver Figura 4) al módulo Gestor de Metadatos, que lo utiliza como referencia temporal a introducir en el metadato. El módulo Gestor de Metadatos añade la información temporal al resto de datos propios del metadato y envía el metadato al módulo de Transmisión de Metadatos que se encarga de encapsularlos y sincronizarlos.

Utilizando la referencia temporal, los metadatos se insertan en el flujo de TS de metadatos dentro de paquetes PES (*Packetized Elementary Stream*) [5] como se muestra en la Tabla I. Así la presencia de los metadatos no perturba a elementos de red intermedios ni de procesamiento del flujo multimedia que pudieran utilizarse.

Tabla I. Inserción de metadatos en paquetes PES

| Sintaxis                 | Nº de bits | Mnemónico |
|--------------------------|------------|-----------|
| PES_packet() {           |            |           |
| packet_start_code_prefix | 24         | bslbf     |
| stream_id                | 8          | uimsbf    |
| PES_packet_length        | 16         | uimsbf    |
| {                        |            |           |
| '10'                     | 2          | bslbf     |
| PES_scrambling_control   | 2          | bslbf     |
| }                        |            |           |

|                           |    |        |
|---------------------------|----|--------|
| PES_priority              | 1  | bslbf  |
| data_alignment_indicator  | 1  | bslbf  |
| copyright                 | 1  | bslbf  |
| original_or_copy          | 1  | bslbf  |
| PTS_DTS_flags             | 2  | bslbf  |
| ESCR_flag                 | 1  | bslbf  |
| ES_rate_flag              | 1  | bslbf  |
| DSM_trick_mode_flag       | 1  | bslbf  |
| additional_copy_info_flag | 1  | bslbf  |
| PES_CRC_flag              | 1  | bslbf  |
| PES_extension_flag        | 1  | bslb   |
| PES_header_data_length    | 8  | uimsbf |
| {                         |    |        |
| '0010'                    | 4  | bslbf  |
| PTS & marker bits         | 36 | bslbf  |
| }                         |    |        |
| } for (i=0;i<N1;i++) {    |    |        |
| stuffing_byte             | 8  | bslbf  |
| } for (i=0;i<N2;i++) {    |    |        |
| PES_packet_data_byte      | 8  | bslbf  |
| }                         |    |        |

### III. APLICACIONES

La versatilidad de la arquitectura aquí presentada permite aplicar el uso de metadatos no sólo en aplicaciones de videoconferencia, como es el caso del CENIT VISION, sino también en diferentes sectores como son: seguridad o vigilancia, en el que se utilizan metadatos de presencia e incluso los que informan del número de personas que se encuentran en un determinado área; la televisión digital, que permite la personalización de los contenidos, ofreciendo distintos programas a distintos usuarios; publicidad personalizada, que se ajuste a los gustos e intereses de los usuarios; etc.

Por otra parte, los terminales de usuario integran cada vez más capacidades que pueden emplearse para enriquecer la comunicación. Se pueden tomar, por ejemplo, los teléfonos móviles de última generación, que integran videocámara, geolocalización, acelerómetro, brújula, sistemas empotrados de reconocimiento del habla... Existe un mercado emergente para la creación de nuevas aplicaciones basadas en la integración de todas estas capacidades. Así pues, resulta natural plantearse un entorno en el que esta información adicional pueda emplearse para enriquecer la comunicación.

#### A. Aplicación en Videoconferencia 3D: proyecto CENIT VISION

En el proyecto CENIT VISION se ha desarrollado la arquitectura descrita en este documento. Los metadatos se generan durante el transcurso de las videoconferencias y son transmitidos por un canal independiente. Al estar sincronizados con el flujo multimedia, estos metadatos posibilitan la integración de numerosos servicios de valor añadido con la plataforma de Videoconferencia 3D.

Dentro de los servicios habilitados por los metadatos sincronizados se distingue entre: aquellos que hacen uso de los metadatos durante las propias sesiones de videoconferencia (servicios *online*) y aquellos que procesan los metadatos a posteriori (servicios *offline*).

Los servicios *online* consumen los metadatos según se van generando y transmitiendo. Permiten por lo tanto tomar decisiones y ejecutar aplicaciones durante la propia ejecución del servicio de videoconferencia. Ejemplos de estos servicios son:

- Monitorización del servicio: los metadatos se utilizan para transmitir información acerca de las sesiones establecidas. Los datos enviados describen el estado de la red, el nivel de congestión en la misma, el perfil de los usuarios (p. ej., tipo de terminales con los que cuentan), etc. Esta información es muy útil para la plataforma de Videoconferencia 3D, ya que permite conocer el estado de su infraestructura en cada instante y, consecuentemente, tomar decisiones que garanticen la QoS ofertada y la resolución rápida de las incidencias detectadas.
- Servicios bajo demanda: la plataforma de videoconferencia ofrece a los usuarios la posibilidad de realizar diferentes acciones durante la comunicación. Estas acciones se traducen en metadatos que permiten lanzar nuevos servicios asociados. Ejemplos de estos servicios son: grabación y etiquetado de la sesión, el usuario decide grabar la videoconferencia para su posterior visualización o post-procesado; anotación de eventos, el usuario decide marcar instantes de la sesión que considera de su interés, y que podrán ser empleados posteriormente; juegos interactivos, el usuario realiza movimientos que son analizados en tiempo real y transmitidos en forma de metadatos.
- Servicios automáticos: el análisis de sesiones en tiempo real proporciona información acerca de eventos de diferente naturaleza que se gestiona de forma automática por la Plataforma de Videoconferencia, es decir, sin participación activa por parte de los usuarios. Así, por ejemplo, eventos de presencia (cuándo un nuevo individuo aparece o

abandona la escena), de proximidad (definiendo regiones de seguridad), o de reconocimiento gestual (petición de socorro) respectivamente, habilitan servicios de videovigilancia, servicios de generación de alarmas y servicios de respuesta rápida derivadas de las alarmas detectadas.

Los servicios *offline* consumen los metadatos una vez han finalizado las sesiones de videoconferencia, por lo cual deben capturar la metainformación que está siendo transmitida y almacenarla de forma que sea accesible posteriormente. Estos servicios están orientados sobre todo al análisis del servicio ofrecido, aunque también permiten ofrecer nuevas funcionalidades a los usuarios finales. Ejemplos de este tipo de servicios son:

- Estadísticas: se analizan los metadatos de monitorización de QoS. Información acerca de los paquetes perdidos, la congestión de la red, las interacciones de los usuarios con los servicios bajo demanda, las alarmas generadas, etc. se transmite periódicamente, o bajo demanda, a la Plataforma de Videoconferencia, la cual procesa estos datos conjuntamente con la información recolectada del resto de sesiones establecidas. El análisis de este procesamiento conjunto permite tomar decisiones sobre el servicio de ámbito global, a la vez que desde un punto de vista de negocio aporta conocimiento imprescindible sobre el dimensionamiento de la red, los intereses de los usuarios y sus interacciones con los servicios de valor añadido ofertados en la plataforma.
- Post-procesado de sesiones: las sesiones grabadas son post-procesadas en base a los metadatos asociados capturados. Esta información acerca de eventos relevantes ocurridos durante la sesión habilita múltiples posibilidades de procesamiento desde el punto de vista de los metadatos: clasificación de las sesiones en función de los eventos que contienen, segmentación de los instantes relevantes de la sesión, análisis de las alarmas detectadas, extracción de imágenes a modo de fotografías, etc. Los resultados de este post-procesamiento permiten ofrecer toda una gama de nuevos servicios a los usuarios finales.

En el proyecto VISION, en concreto, se han implementado los siguientes servicios basados en el nuevo flujo de metadatos sincronizados con la videoconferencia 3D: grabación y etiquetado de sesiones a petición de los usuarios, detección de presencia y de proximidad con elementos de la escena, reconocimiento gestual, detección de movimiento, captura/liberación de objetos por parte de los usuarios, generación de resúmenes y clasificación de sesiones. Dichos servicios están integrados en el Piloto desarrollado por el Proyecto CENIT VISION.

### B. Aplicación en Distribución Multimedia e IPTV

Otro campo de aplicación es la distribución de televisión sobre IP (IPTV). En este ámbito, la introducción de metadatos sincronizados con el vídeo permitirá enriquecer también los servicios audiovisuales que se proporcionen.

Además de las aplicaciones mencionadas en el apartado anterior (muchos de ellos inmediatamente aplicables en el campo de la distribución de contenidos multimedia), el caso de uso más relevante a corto plazo es la composición de contenidos personalizados para los usuarios, con especial foco en la publicidad dirigida. Esta aplicación se ha estandarizado recientemente como parte de la arquitectura de IPTV integrada en NGN de TISPAN [6], tomando como base las normas SCTE-35 [7], SCTE-67 [8] y SCTE-130 [9].

La idea principal consiste en marcar determinados fragmentos del flujo de vídeo (oportunidades) para que puedan ser reemplazados por otros (normalmente anuncios publicitarios) en algún punto de la cadena de transmisión (puede ser tanto en un sistema de red intermedio como en el terminal de usuario). Para conseguir esto es necesario señalar con precisión los instantes iniciales (punto de entrada) y finales (punto de salida) del fragmento, así como dar información adicional (metadatos) sobre el contenido del mismo [7]. La Figura 5 muestra el diagrama de bloques de un sistema de reemplazo de vídeo (*Video Splicing*).

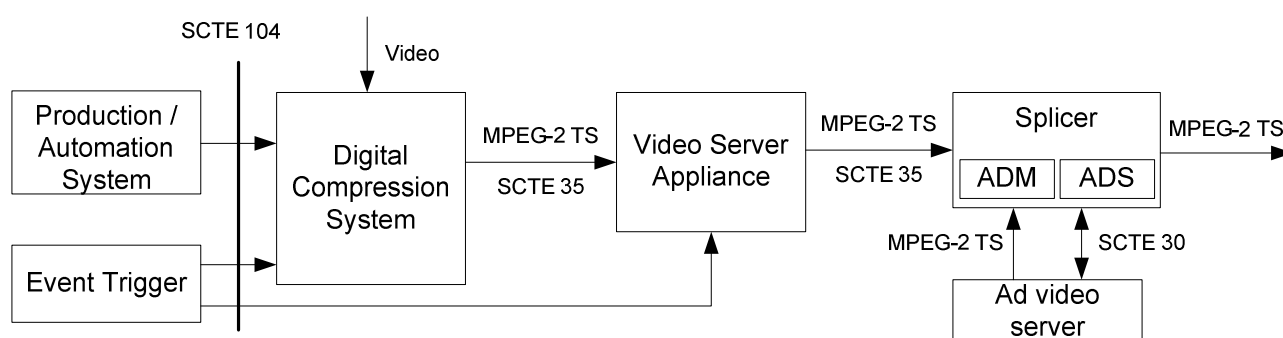


Figura 5. Sistema genérico de Video Splicing y flujo de mensajes SCTE 104, SCTE 35 y SCTE 30.

Las oportunidades de reemplazo pueden ser planificadas previamente por un módulo de gestión (*Production/Automation System* en la Figura 5) o identificadas dinámicamente (lanzadas por un módulo *Event Trigger*). Ambos módulos comunicarían, cumpliendo el API definido en la norma SCTE 104 [10], las oportunidades de reemplazo a los módulos encargados de insertar estas oportunidades en el flujo.

Las oportunidades insertadas en el flujo se señalizan en un PID del TS según define la norma SCTE-35 (*SCTE 35 Cue PID*), y pueden ser insertadas tanto por codificadores de vídeo (*Digital Compression System* en la Figura 5) como por equipos de red intermedios de procesamiento de vídeo (*Video Server Appliance*).

Cuando el módulo de reemplazo de vídeo (*Splicer* en la Figura 5) detecta una oportunidad (un mensaje SCTE-35) puede reemplazar el contenido actual por otro (p. ej., un anuncio personalizado), que es elegido por un servicio de toma de decisiones (ADS, *Ad Decision Service*) en función del contenido y del perfil de usuario objetivo, y que puede ser determinado mediante distintas técnicas (p. ej., en función de las costumbres de visualización, tras una selección directa por parte del usuario...). Las decisiones tomadas por un ADS pueden ser muy sencillas (p. ej., poner un anuncio a una hora en un determinado contenido) o muy sofisticadas (p. ej., basadas en datos de los usuarios, zona donde se emite el anuncio, etc.). Las políticas de funcionamiento ADS se establecen desde el ADM (*Ad Management Service*), que gestiona las oportunidades de colocar contenidos de reemplazo a partir de la información recibida de otros servicios de la arquitectura [6].

El módulo *Splicer* solicita el contenido de reemplazo a un servidor de vídeo (*Ad Video Server* en la Figura 5) por medio del API definido en la norma SCTE-30 [11].

#### IV. CONCLUSIONES

La arquitectura presentada permite insertar y sincronizar de forma eficiente metadatos dentro de una cadena de transmisión multimedia compleja, sin afectar a la comunicación multimedia en sí y manteniendo tanto el flujo de información multimedia como el flujo de metadatos asociados sincronizados en el sistema, de modo que puedan tratarse y procesarse por módulos intermedios de red o aplicaciones terminales específicas.

De esta manera, los metadatos no interfieren con el flujo multimedia principal en lo que se refiere a la decodificación de audio y vídeo, por lo que se comportan de forma transparente para un terminal de usuario receptor que no vaya a utilizar la información asociada, reduciendo el impacto en los costes de integración.

Este sistema habilita la provisión de servicios de valor añadido basados en metadatos. Debido a que el sistema permite asociar con precisión la metainformación a instantes o períodos de tiempo dentro del flujo multimedia y, además, transmite ambos flujos en paralelo, estos servicios de valor añadido pueden localizarse en cualquier nodo de red que procese el contenido y pueden funcionar de forma indistinta en tiempo real o en diferido. Estas características permiten desarrollar un amplio abanico de servicios relacionados con las comunicaciones multimedia. En este trabajo se han descrito alguno de ellos, así como la aplicabilidad de esta arquitectura en otros entornos como la videovigilancia.

#### AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el CDTI, en el marco del proyecto CENIT-VISION 2007-1007.

#### REFERENCIAS

- [1] Anne. R. Kenney, Oya Y. Rieger, Richard Entlich, "Tutorial de Digitalización de Imágenes", Universidad de Cornell, <http://www.library.cornell.edu/preservation/tutorialspanish/metadata/table5-1.html>. 2003.
- [2] ISO/IEC 15938. "Information Technology – Multimedia Content Description Interface". MPEG-7. 2003.
- [3] José M. Martínez, "MPEG-7 Overview", ISO/IEC JTC1/SC29/WG11N6828, Oct. 2004.
- [4] Pérez, L., Rizardos, J., Boutanaach, A., Palacios, J.M., Landabaso, J.L. and Cano, F., "VISIÓN: Comunicaciones Audiovisuales de Nueva Generación", Telecom I+D, Bilbao (España), ISBN: 978-84-9860-135-0. 2008.
- [5] ISO/IEC 13818-1. "Information Technology – Generic Coding of Moving Pictures and Associated Audio: Systems". MPEG2-Systems International Standard. Nov. 1994.
- [6] TISPAN ETSI TS 182 028 V3.3.1 "Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN). NGN integrated IPTV subsystem Architecture". 2009-10.
- [7] SCTE-35: "Digital Program Insertion Cueing Message for Cable". 2007.
- [8] SCTE-67: "Digital Program Insertion Cueing Message for Cable – Interpretation for SCTE 35". 2006.
- [9] SCTE-130: "Digital Program Insertion - Advertising Systems Interfaces". 2008.
- [10] SCTE-104: "Automation System to Compression System Communications Applications Program Interface (API)". 2004.
- [11] SCTE-30: "Digital Program Insertion Splicing API". 2009.